

## LONG-TERM PRESERVATION

Government entities generate large amounts of electronic records in various formats. Many of these records are useful only for a short period of time, but others are considered archival and thus need to be kept permanently. To ensure the reliability of these records with continuing value over time, establish and implement a preservation plan. Since electronic records create access challenges, preservation plans for electronic records must consider the limitations of storage media, the probability of hardware and software obsolescence, and the ways in which information may potentially be used.

When developing a preservation plan, first conduct a needs-analysis to help guide decision-making. How records are used will aid in determining appropriate preservation options as well as the types of metadata that will be most useful to create with the records.

Decide whether a record must be kept in electronic format or whether there is another cost-effective option for long-term storage. For example, a word processing document could be printed on paper, which might take up more physical space but would not require further migration or conversion. Printing a copy of a Web site, however, would result in the loss of the majority of its functionality. If an agency or locality chooses to retain custody of permanent electronic records, the agency or locality assumes responsibility for maintaining their reliability, authenticity, integrity, and usability.

It is also important to ascertain if access to certain data in the records is restricted by statute, such as the [Freedom of Information Act](#) (FOIA) and other state and federal laws. Long-term storage and access policies must address these obligations.

### Digital Preservation Techniques

[Preservation](#) plans for electronic records must consider the probability of hardware and software obsolescence and guarantee long-term access to records. Proprietary software will eventually become obsolete as companies upgrade or stop producing the product altogether. There are several approaches, some more practical than others, to ensure that electronic records remain useful over time.

One approach is **emulation**. Emulator programs simulate the behavior, look, and feel of other programs, thus preserving the functionality of the records in their original format without the necessity of saving the original equipment and software. Emulation, however, has so far proved

more attractive in theory than in practice. There are few examples of success using this approach, and costs have been shown to be high. It has a further limitation in that, at best, emulation simply reproduces earlier, less-sophisticated versions of an application.

Another approach to preservation is **encapsulation**. It involves combining the object to be preserved with all of the necessary details on how to interpret it within a wrapper or package, all possibly formatted in XML. While appealing in its comprehensiveness, encapsulation has several drawbacks: file sizes are large because of all of the included information; format specifications must be determined; the encapsulated records must somehow be generated, usually separate from the act of record creation; and the encapsulated records must still be migrated over time.

The most common approach to preserving electronic records involves a combination of two other techniques: migration and conversion. **Migration** is the process of moving files to new media (also known as "media refreshing") or computer platforms in order to maintain their value.

**Conversion** entails changing files from one format to another and may involve moving from a proprietary format, such as Microsoft Word, to a nonproprietary one such as a plain text file or XML. To avoid losing data in the process, testing and analysis should determine exactly what changes will occur and whether they are acceptable. With both migration and conversion, special attention must be paid to maintaining the accessibility of associated metadata. When properly planned and executed, the migration and conversion approaches represent the easiest and most cost-effective preservation methods available today. A discussion of file format and media options to consider when migrating or converting records follows this section.

One challenge of converting files is that data may be lost. Data compression is the process of encoding information using fewer bits. Compression saves storage space and enables data to be transmitted quickly and easily, but data may be lost as a result. Compression also introduces an additional layer of software dependency. Compression options vary in their degree of data loss. Some are intentionally "lossy," such as the JPEG format, which relies on the human eye to fill in the missing detail. Others are designed to be "lossless."

There are three basic types of loss that may occur during conversion or migration:

- *Data*. If data is lost, the content of the record is lost to a varying degree.
- *Appearance*. Converting records may alter the formatting of the file. For example, converting all word processing documents to RTF may cause some loss of page layout. Determine whether this loss affects the completeness of the record. If the structure is essential to understanding the record, this loss may be unacceptable.

- *Relationships*. Relationships within the data in a file, such as spreadsheet cell formulas and database file fields, may be lost.